

REMARKS

This is in response to the Office Action mailed on June 1, 2006. In the Office Action, claims 1-24 were pending.

The Office Action first reports that claims 1-24 were subject to a restriction requirement. In particular, restriction was required to an invention in group I (claims 1-8) or group II (claims 9-24). Applicants hereby elect to prosecute claims 1-8 and have cancelled claims 9-24 with this amendment.

The disclosure was objected to for containing an embedded hyperlink on page 15. With this amendment, the hyperlink has been removed. Withdrawal of the objection is thus requested.

Claims 1-8 were rejected under 35 U.S.C. 101 for being directed to non-statutory subject matter. In particular, the Office Action reported that the claims did not recite any limitation wherein a tangible and concrete result would inherently flow from the claimed invention. With this amendment, independent claim 1 has been amended to recite a step of "outputting". Similarly, independent claim 5 has been amended to recite that an extraction module provides an "output". Applicants submit that the step of outputting and providing output are directed toward a tangible and concrete result. In particular, a first and a second set of elements are output that can be used for further processing in information extraction. As known to those skilled in the art, extracted information that is output can be used in many different tangible and concrete ways such as rendering on a computer, provided to another module for analysis, etc. Thus, it is believed that claims 1-8 are in compliance with the requirements of 35 U.S.C. 101 and withdrawal of this rejection is respectfully requested.

Claims 1-8 were also rejected under 35 U.S.C. 103(a) as being unpatentable over Nicholas Jr. (2000) (Nicholas hereafter) taken with Eskin et al. (April 2003) (Eskin hereafter). Both Nicholas and Eskin relate to database searching of sequences of proteins. As discussed in Nichols, "The objective of a database search is to distinguish sequences related to the query sequence by some model (e.g., evolution) from unrelated sequences" (see page 1174, column 1). The query sequence is thus compared to other sequences to determine a similarity between the

query sequence and the other sequences. As discussed in Eskin, this can be used to “automatically classify unknown proteins into families” (see page 187).

In contrast, subject matter disclosed in the present application is directed to information extraction from a plurality of documents. As discussed in the application, extracting information from a source is performed to output related elements pertaining to a topic. For example, a company/product pair can be extracted from documents that are related to a product release. The extraction is performed to relate the information to a general topic. This situation is thus different from a protein database search that merely finds a similar sequence of proteins.

In view of the differences between database search and information extraction, applicants have amended independent claims 1 and 5 to clarify the features recited therein. Claim 1 has been amended to recite a computer-implemented method of extracting information from an information source comprising a plurality of documents. The method includes accessing strings of text in the information source and comparing the strings of text in the information source with generalized extraction patterns. A plurality of strings in the information source are identified that match at least one generalized extraction pattern. The generalized extraction patterns include words and wildcards, wherein the wildcards denote that at least one word in an individual string can be skipped in order to match the individual string to an individual generalized extraction pattern. The method also includes extracting a first set of related elements of text pertaining to a topic from a first string of the plurality of strings based on a corresponding set of related elements pertaining to the topic in the at least one generalized extraction pattern. The first string is associated with a first document in the plurality of documents. The method also includes extracting a second set of related elements of text pertaining to the topic from a second string of the plurality of strings based on the corresponding set of related elements in the at least one generalized extraction pattern. The second string is associated with a second document in the plurality of documents. At least one of the related elements of text in the first set of related elements is different from each of the related elements of text in the second set of related elements of text. The first related set of elements and the second set of related elements are output.

Similarly, independent claim 5 has been amended to recite a computer-readable medium for extracting information from an information source comprising a plurality of documents. The medium includes a data structure including a set of generalized extraction patterns including words and an indication of a position for at least one optional word. An extraction module uses the set of generalized extraction patterns to match a first string and a second string in the information source with one of the generalized extraction patterns. The first string is associated with a first document in the plurality of documents and the second string is associated with a second document in the plurality of documents. The extraction module also extracts a first set of related elements of text pertaining to a topic from the first string based on a corresponding set of related elements in said one of the generalized extraction patterns and a second set of related elements of text pertaining to the topic from the second string based on the corresponding set of related elements in said one of the generalized extraction patterns. At least one of the related elements of text in the first set of related elements is different from each of the related elements of text in the second set of related elements of text. The extraction module also outputs the first related set of elements and the second related set of elements.

Features recited in claims 1 and 5 are neither taught or suggested by the combination of Nichols and Eskin. First, the features in the claims relate to extracting information from a plurality of documents where Nichols and Eskin perform a database search on sequences of proteins. Furthermore, the features in the claims relate to extracting first and second related sets of text from first and second documents. While Nichols and Eskin may describe comparing a query to sequences in a database, there is no mention of extraction or extracting sets of related information from an information source. If a high similarity is found between a query and a sequence in the database, the query is classified to a particular family. There is no teaching or suggestion that information from the sequence (or from multiple sequences) is extracted that pertains to a particular topic. Furthermore, there is no teaching or suggestion that a first set of extracted elements differ from each of the extracted elements from a second set of related elements.

To illustrate an example for utilizing the features of claims 1 and 5, a plurality of news articles could be a plurality of documents. A topic could relate to product release information and a set of related elements could be a company and a product. Features recited in claims 1 and 5 relate to matching strings in the news articles and extracting a first and second set of elements. For example, the first set of elements could include a first specific company and a first specific product. At least one of the first specific company and the first specific product is different from elements in the second set of related elements. There is simply no teaching or suggestion in Nicholas and Eskin for extracting different sets of elements from different sequences related to a general pattern. Instead, the exact proteins are matched to exact proteins. As a result, independent claims 1 and 5 are believed to be allowable. Additionally, claims 2-4 and 6-8 are believed to be allowable at least on their relation to claims 1 and 5, respectively.

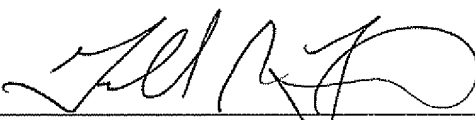
Applicants have further added claims 25-29, which depend from claim 1, and claims 30-34, which depend from claim 5. These claims recite features related to information related to the example discussed above and are also believed to be allowable.

In view of the foregoing, Applicants submit that the present application is in condition for allowance. Withdrawal of the rejections and allowance of the pending claims is respectfully requested.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: 
Todd R. Fronek, Reg. No. 48,516
900 Second Avenue South, Suite 1400
Minneapolis, Minnesota 55402-3319
Phone: (612) 334-3222 Fax: (612) 334-3312

TRF:rdg